



# Australian Bureau of Statistics

## **1352.0.55.051 - Research Paper: Winsorization for generalised regression estimation (Methodology Advisory Committee), Nov 2002**

Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 20/05/2002 First Issue

---

## Summary

### About this Release

#### About this Release

The availability of Business Activity Statement (BAS) data collected by the Australian Taxation Office (ATO) has provided the Australian Bureau of Statistics (ABS) with opportunities to improve the efficiency of sample design and estimation for its business surveys. ABS business surveys currently use two methods of estimation; number-raised estimation and ratio estimation. While ratio estimation allows the use of one auxiliary variable to improve the precision of the estimates, generalised regression (GREG) estimation allows the use of more than one auxiliary variable, and hence has the potential to be more efficient (i.e. reduce the current sample sizes for ABS business surveys with no reduction in the precision of the estimates) than number-raised and ratio estimation.

The generalised regression estimator is unbiased with respect to the assumed model. However, if by chance there are several units in the sample with unusually large residuals under the generalised regression model, then the generalised regression estimator may grossly underestimate or overestimate the population totals. One solution to this problem is to modify values outside preset cutoff values to values closer to these cutoff values. This estimator is called the "winsorized" estimator. Although the winsorized estimator is biased, it may have a considerably smaller mean squared error than the generalised regression estimator.

There often exist linear relationships between the various data items collected and derived in ABS business surveys, and it is important that these linear relationships still hold after winsorization. The current ABS estimation system allows the linear relationships to be maintained by two methods. Unfortunately, there are some situations where these two methods perform quite poorly. An alternative method which attempts to overcome the shortcomings of the two methods is suggested, which requires the specification of a distance function between the original and final winsorized values. Although any one of a number of distance functions could be used, the one examined in this paper is the generalised least squares distance function.